Journal of Surgical Oncology 1997;66:151–155

GUEST EDITORIAL

Surveillance After Potentially Curative Cancer Treatment

FRANK E. JOHNSON, MD*
Department of Surgery, St. Louis University Health Sciences Center, St. Louis, Missouri

Until the twentieth century, cure of cancer was rare and cancer patient follow-up was concerned with palliative care delivered to dying, symptomatic patients. With modern surgery, cures became possible with regularity for the first time in history [1]. This created a population with unique attributes—they had survived treatment carried out with curative intent, had a realistic prospect of having been cured, and were more or less asymptomatic. The natural history of this then-novel patient population soon indicated that these patients were more vulnerable than the population at large to certain types of cancer [2].

Although instances of multiple primary cancers of a single organ were generally considered medical curiosities and reported as single instances until well into the twentieth century, inklings of genetic predisposition to cancer began to occur by this time. Nonetheless, major medical textbooks written earlier in this century make little mention of the phenomenon of second primary neoplasms in patients treated for common carcinomas—breast, upper aerodigestive tract, large intestine, etc. Many common cancers, and a number of uncommon ones as well, are now known to be markers for synchronous or metachronous new carcinomas in the same or other organs, giving rise to the concept of organ-specific surveillance. A good example of this is breast cancer. It was realized long ago that cure of cancer in one breast could be followed by the development of cancer in the contralateral breast [3], and that the second cancer also could be treated with curative potential. Thus the detection of new organ-specific cancers became enshrined in clinical practice.

Initiatives to develop consensus are becoming more important as information about medical matters proliferates. It is clear that surveillance strategies for patients being followed after potentially curative treatment often vary widely among experts, that the cost of surveillance can vary greatly [4,5], and that the effectiveness of the various strategies has been poorly documented in almost every instance.

Two goals, detection of recurrence of the index lesion and detection of second primary tumors, have been the foundation of surveillance strategies. Patients with recurrent cancers such as testicular carcinoma or Hodgkin's disease often can be cured, and even those with incurable recurrences of most sorts of cancer often can receive effective palliation. With certain other tumors, unfortunately, presymptomatic detection of recurrence is currently of little practical value because effective therapy is not available, which calls into question whether active surveillance is beneficial.

A third goal is documentation of the results of therapy. This has been aided by improvements in medical record documentation. It is instructive to read today the fragmentary histories, physical examinations, and diagnostic tests upon which important clinical decisions were made in the past, and it is not surprising that treatment outcomes were difficult to predict and often poor. Medical recordkeeping is an underappreciated endeavor that has greatly improved our ability to analyze the outcomes of clinical patient care. Retrospective reviews of patient records have led to continuous improvements in patient care and have improved our ability to discern the future for patients with cancer. This, in turn, has helped shape follow-up strategies after cancer therapy.

A fourth goal of postoperative surveillance also is undoubtedly psychological. There is little doubt that some patients derive comfort from receiving a clean bill of health from their doctor after a clinic visit, since fear of recurrent cancer is so common, and so often well-warranted. There is little written about the psychological benefits to the patient of receiving posttreatment surveillance, although one recent well-controlled trial of surveillance after breast cancer treatment did not find an impact on quality of life resulting from intensive surveil-

*Correspondence to: Frank E. Johnson, M.D., Department of Surgery, St. Louis University Health Sciences Center, 3635 Vista Avenue, St. Louis, MO 63110-0250. Telephone: 314-577-8316; Fax: 314-771-1945.

Accepted 4 February 1997

lance when compared to a nonintensive strategy [6]. It must be mentioned that there are psychological disadvantages of such surveillance as well. Most clinicians are only dimly aware how their patients feel about undergoing the battery of tests that comprise surveillance, since the feared recurrence may be detected by any of the surveillance tests utilized. The optimal management of a patient who has undergone potentially curative treatment for cancer requires considerable clinical wisdom. This necessarily must synthesize knowledge, insight into patient's concerns [7], and, to an ever-larger extent, maneuvering within externally imposed guidelines.

A secondary gain inherent in any surveillance strategy, and one that is not commonly taken into account in assessing the utility of follow-up management plans, is the detection of other conditions that may be of medical significance [4]. Comprehensive accounting of the value of surveillance should consider unanticipated benefits such as these.

The reliable possibility of cure after cancer therapy represented a milestone event in medicine and one that has proved to have important socioeconomic implications. This very positive outcome was strongly desired, in spite of expense, by doctors and patients. The success of surgery was shortly followed by the development of ionizing radiation as a reliable treatment tool. Effective systemic therapy has been available for a century but largely consisted of hormonal ablative treatments until the introduction of toxic drugs, such as nitrogen mustard, into clinical medicine after World War II [8]. Although systemic therapy for the common solid tumors made relatively scant progress until recently, chemotherapy for upper aerodigestive tract carcinoma, breast carcinoma, large bowel carcinoma, and other common cancers is now incorporated into standard care. A fraction of patients receiving chemotherapy are long-term, disease-free survivors in unmaintained remission, now conceptualized as cured patients. Surgery itself can be considered adjuvant therapy in organ-conserving treatment of breast cancer. How adjuvant therapy should affect posttreatment surveillance strategies is almost completely unknown. This subset of patients forms a population in which screening for new primary cancers or recurrence of the index neoplasm has been vigorously pursued. The screening of these patients has been aided by ever more powerful diagnostic tests.

The profusion of tests and their obvious utility in many clinical situations has led to a dilemma. How can clinicians best select which tests to order in their patients who have received treatment for cancer and who are now seemingly well? How often should tests be ordered? Which ones are likely to provide useful new information and which are not? How does the information affect clinical decision making and patient outcomes? Clinicians now employ these tests in their patients treated with curative intent for cancer, usually without rigorous prior demonstration of the usefulness of the tests and often with little knowledge of the costs involved.

For example, many factors are now known to indicate an increasing risk of treatment failure, even if initial therapy was administered with curative intent. Thus there is an apparent rationale favoring increased surveillance in such patients. Whether more intensive surveillance pays off in terms of lives saved or improved quality of life has been investigated rather little, however. Russell [9] has provided a scholarly analysis of this issue for carcinoma of the cervix, a neoplasm for which the Papanicolaou (Pap) test has been well established as an effective screening tool. Her analysis concerns primary screening of patients without prior cancer, but the message extends to cancer patient follow-up quite nicely. The value of surveillance for cervical cancer is indisputable, but how often should screening be carried out? A large impact on the death rate from cervix cancer is achieved by screening adult women every 10 years. This is improved by more frequent screening, but at ever-increasing marginal cost and ever-decreasing marginal gain in life expectancy. For example, Pap testing every 3 years is estimated to increase mean life expectancy by only 1.5 days as compared to testing every 4 years. This leads to a consideration of costs. Patients, insurance companies, health care policy makers, and physicians all view this issue somewhat differently, for obvious reasons. Assuming a cost of $75 per Pap test, screening every 3 years for every woman from 20–74 years old in the United States would cost about $2 billion per year, whereas annual screening would cost $6 billion per year. Such sums of money warrant scrutiny. Furthermore, surveillance testing provides a diagnosis only, and costs of active treatment must be factored into any rational analysis. For cervix cancer, Eddy [4] calculates that Pap testing every 4 years saves 1 year of life at a cost of about $10,000, whereas annual screening costs about $40,000 per year of life saved, as compared to a policy of no screening. However, when annual surveillance is compared with surveillance every 2 years, the cost per additional year of life saved by the more intensive strategy exceeds $1 million. Such considerations are causing changes in health care policy. Society, if it behaves rationally, should not neglect the small risk that detectable and curable cancer may exist in certain patient groups, but should avoid excessive diagnostic testing and treatment. Society cannot rationally afford to squander its resources in low-benefit, high-cost enterprises, particularly when these also disrupt patients' lives needlessly and impose their own risks.

Once the causal role of environmental and genetic factors in carcinogenesis became clear, preventive measures were integrated into posttreatment follow-up efforts. For decades, physicians have advised lung cancer patients

during their follow-up visits to stop smoking. Skin cancer patients are cautioned to minimize sun exposure, and so on. We now know that diet, obesity, and other modifiable risk factors exist for certain cancers and we can counsel our patients appropriately. Active therapy in the form of many different sorts of drugs (retinoids, tamoxifen, finasteride, etc.) are now being tested in clinical trials for their activity in preventing common cancers [8]. If these chemopreventive regimens are found to be effective, monitoring of intermediate endpoints and drug toxicities will provide new justification for follow-up in large populations of cancer patients. How follow-up strategies should be modified to accommodate chemoprevention regimens is not known.

We are increasingly being forced to realize that the dramatic advances in diagnostic testing have implications for clinical medicine and for medical economics [4,10]. CT scans, sonograms, and other modern imaging tests can provide remarkable anatomic detail. Deeply seated lesions can be detected before they produce any clinical disease. It is common for abnormalities in serological tumor markers, endoscopy, and the like to antedate clinical disease, but all these tests are costly. The profusion of medical information also can make clinical decision making ambiguous. What is the physician to do, for example, when a breast cancer patient given potentially curative treatment several years earlier develops an isolated abnormality on bone scan? Further expensive tests may well be indicated and biopsy proof of recurrent cancer may be difficult to obtain. Toxic therapy might be given needlessly if the breast cancer had not actually recurred. Crippling complications could develop if cancer had recurred, but therapy was withheld.

Thus, paradoxically, advances in diagnosis can make the doctor's work more difficult and the patient's decision making more confusing. Some of the confusion dwells in documenting the anatomical distribution of cancer and defining the value of cancer treatment. The images produced by the modern diagnostic armamentarium can be spectacular and compelling. A sufficiently abnormal image may prompt a flurry of additional diagnostic and therapeutic interventions that sometimes yield few benefits and in fact may be detrimental. The apparent incidence and prevalence of a disease increase as our ability to detect the disease improves [10]. In the case of tumor marker substances in blood or other body fluids, abnormalities are loosely related to tumor size. For diagnostic imaging tests, the size of the radiographic abnormality correlates highly with diagnostic sensitivity and specificity. With improved imaging tests, smaller and smaller abnormalities can be reliably detected. A cycle may develop in which some form of new imaging technique decreases the threshold for detection of cancer, which yields a higher percentage of positive diagnoses in a given population. This is taken as providing an improvement in clinical management, which reinforces the increased use of the imaging test. These considerations are important when deciding how to follow cancer patients after primary treatment.

The assessment of diagnostic accuracy, which rests on concepts such as sensitivity, specificity, and the like, uses short-term pathological interpretation to validate these tests. This assessment commonly incorporates neither the long-term clinical implications of the diagnosis nor the effects of therapies. Addressing these considerations requires different sorts of evidence, which are usually more difficult to come by. Nonetheless, more sensitive tests usually replace less sensitive ones, even though some of the cancers detected by very sensitive tests are of unknown natural history and have an unknown response to treatment. An example of this is cervical intraepithelial neoplasia, the detection of which is greatly aided by the Pap test. Frequent Pap testing established this diagnosis in many women and led them to receive hysterectomy, until it became clear that equivalent survival could be achieved with lesser ablative procedures.

This simple example shows how shifting perceptions of cancer prevalence and unclear documentation of therapeutic effectiveness can prompt increased diagnostic testing and excessive treatment [10]. It illustrates how improved diagnostic tests may not help the clinician reach a decision about how the newly detectable subclinical disease should best be treated. Because the incidence and prevalence of detected disease increase as the sensitivity of testing methods increase, patient outcomes often appear to improve, because of noncomparability with historical controls. This seeming gain in treatment outcome further reinforces the initial trend in favor of an intensive testing strategy prompted by improved test sensitivity. As the example of Pap testing shows, this can lead to increased treatment intensity and even greater use of the diagnostic tests. In addition, the establishment of a cancer diagnosis is not always accompanied by the development of clinical disease [11] and certainly does not establish whether a patient will benefit from a therapeutic intervention. Clinical trials can help avoid these problems. Large-scale clinical trials of screening in high-risk groups for breast cancer, lung cancer, and the like have often yielded results not favoring frequent testing.

Cancers are currently defined in terms of microscopic morphologic criteria and thus the prevalence of cancer increases with our ability to sample large amounts of tissue. Increasingly sophisticated imaging techniques can be so sensitive as to call into question what is meant by the term ''cancer'' [11]. An excellent example of this is found in the field of breast cancer. Before mammography, breast cancer was generally diagnosed clinically because a breast mass was discovered. However, breast cancers detected by screening mammography are often so small as to be impalpable. In addition, they often are

in situ lesions, the natural history of which is not very clear. The ever-lower threshold for breast cancer detection helps explain the increased prevalence of cancer on mammographic screening and the increase in the percentage of carcinomas in situ among breast cancers detected via mammography. This illustrates the fact, relatively unappreciated until recently, that clinically occult cancer is present in many apparently healthy citizens [10–12]. What should clinicians do about such abnormalities that are detected during surveillance of patients who have been treated for cancer?

The advances in surveillance are thus changing our notions about the prevalence of cancer. They also alter perceptions of the natural history of disease and the response of cancer to treatment. Unfortunately, in selecting tests and treatments to offer our patients, we are rarely able to rely on the results of well-controlled prospective clinical trials, which are expensive and difficult to carry out. Thus physicians must usually rely on less rigorous sources of guidance such as consensus-development panels, traditions imparted during medical school and residency training, and knowledge of nonscientific issues, such as patient preference and the constraints of third-party payers.

Common sources of bias in deciding on the value of diagnostic and therapeutic interventions are lead time and length bias [10]. Some of the confusion resulting from improvements in diagnostic testing will undoubtedly persist, since medicine remains an art as well as a science. However, some confusion can be eliminated if efforts can be taken to think through the implications of diagnosis of tiny neoplasms of uncertain significance. Stratification of diagnostic certainty according to tumor size and adjusting for the sensitivity and specificity of a given method of detection could make estimation of disease prevalence more independent of the continually evolving imaging methodologies. The reliability of historical comparisons used to assess the utility of newly introduced tests and treatment strategies could be improved by minimizing lead time and length bias, but this is easier said than done. The incidence and prevalence of many diseases can only be expected to increase in the future as imaging modalities, molecular genetic tools, etc., continue to sharpen their focus. Some of these increases can be predicted [11]. We must restrain our temptation to treat diseases detected via cancer patient follow-up schemes in novel ways with potentially harmful and expensive therapies. We must also expect that improved diagnostic techniques will be accompanied by seeming improvements in patient outcomes due to lead time and length biases, stage migration, and the like. Because of these and other complexities, well-controlled, randomized clinical trials will be needed more than ever in the coming decades. The insights from such trials are likely to be great, and the time, effort, and money expended should be rewarded with improved patient management.

Deciding to implement a posttreatment surveillance program of any sort ideally should be founded on a judgment that its benefits and savings exceed its risks and costs. The information needed to arrive at this judgment, unfortunately, is difficult to obtain. Cost containment has become a pervasive issue, particularly where access to sophisticated and expensive follow-up modalities can impose a crushing financial burden on society, individuals, corporations, and insurers. Three methods of assessing relationships among costs, savings, risks, and benefits have been particularly influential, all requiring interdisciplinary research. Cost-effectiveness analysis focuses on a measure of effect, such as survival duration, and attempts to describe the costs, savings, and risks of a specific intervention. Included in the analysis is an estimate of opportunities forfeited because certain resources consumed by the surveillance scheme are not available for other uses. Cost utility analysis is an extension of this technique used to factor in several measures of effect simultaneously. The several measures are given relative weights in order to arrive at a global measure of the intervention. This is called the utility. The best-known utility measure is the quality-adjusted life year (QALY), a simultaneous measure of duration of life and quality of life. The cost-benefit analysis reckons the costs and utility of an intervention, derived from a cost-utility analysis, by attaching a price value to the unit of utility. Other considerations such as legal and ethical ones also can be factored in, with attendant increases in complexity; this is often termed a medical technology assessment.

Genetic testing can now reliably predict the development of certain types of cancer. Testing for relevant oncogenes and tumor suppressor genes in peripheral blood leukocytes, feces, sputum, pancreatic juice, etc., can already quantify future cancer risk or diagnose subclinical cancer. Textbooks outlining gene-level diagnostic tests in clinical medicine have begun to appear. Genetic testing has entered commercial use [13], and there seems little doubt that others will rapidly enter clinical practice once they are shown to be sensitive, specific, and affordable. They will likely be rapidly embraced by doctors and patients since, in most cases, obtaining diagnostic specimens is quite safe, comfortable, and convenient. This will predictably result in dilemmas in patient care, as discussed earlier. What will occur when a patient being followed after lung cancer surgery turns up with molecular evidence of cancer in DNA in his sputum, for example, but no lesions on chest X-ray or bronchoscopy? Similar problems were encountered in earlier surveillance efforts like the Lung Cancer Detection Demonstration Project when sputum cytology was positive but chest X-ray was unrevealing. What about the traditional follow-up tests? Will the revolution in molecular medicine

render them irrelevant? Will gene-level treatments emerge with curative potential for advanced cancer? Such developments seem quite possible and would affect surveillance strategies markedly [14].

This niche in medical care, like all other aspects of the human condition, will likely prove impossible to fully rationalize. Help is on the way from several quarters, however. Practice guidelines being fashioned by medical organizations—often as a counterweight to guidelines used by for-profit corporations—will tend to decrease variability among practitioners. Well-done clinical trials have been carried out [6], and more are in the pipeline. Access to huge computer datasets may allow us to figure out what current practices really are. This will provide the substrate for cost-benefit analyses. As data become more accessible and as analytic approaches become more sophisticated, a flood of articles debunking or supporting various clinical practices will likely appear and influence our patient management accordingly. Until then, the practicing clinician, tugged and battered by many countervailing forces, will have to accept uncertainty, keep the welfare of his or her patients uppermost in priority, and be prepared to modify patient management wisely as new knowledge appears.

REFERENCES

1. Wangensteen OH, Wangensteen SD: ''The Rise of Surgery From Empiric Craft to Scientific Discipline.'' Minneapolis: University of Minnesota Press, 1978, p 567–572.
2. Marchant FE, Lowry LD, Moffitt JJ, Sabbagh R: Current national trends in the posttreatment follow-up of patients with squamous cell carcinoma of the head and neck. Am J Otol 1993;14:88–93.
3. Ewing J: Epithelial and other tumors of the breast. In Ewing J (ed): ''Neoplastic Diseases,'' 3rd ed. Philadelphia: W.B. Saunders, 1928, p 546.
4. Eddy DM: Screening for cervical cancer. Ann Int Med 1990;113: 214–226.
5. Virgo KS, Vernava AM, Longo WE, McKirgan LW, Johnson FE: Cost of patient follow-up after potentially curative colon cancer treatment. J Am Med Assn 1995;273:1837–1841.
6. GIVIO investigators: Impact of follow-up testing on survival and health-related quality of life in breast cancer patients. J Am Med Assn 1994;271:1587–1592.
7. Kassirer JP: Incorporating patients' preferences into medical decisions. New Engl J Med 1994;330:1895–1896.
8. De Vita VT: The evolution of therapeutic research in cancer. New Engl J Med 1978;298:907–910.
9. Russell LB: Screening for cervical cancer. In: ''Educated Guesses: Making Policy About Medical Screening Tests.'' Los Angeles: University of California Press, 1994, p 6–24.
10. Black WD, Welch HG: Advances in diagnostic imaging and overestimations of disease prevalence and the benefits of therapy. N Engl J Med 1993;328:1237–1243.
11. Harach HR, Franssila KO, Wasenius V: Occult papillary carcinoma of the thyroid: A ''normal'' finding in Finland: A systematic autopsy study. Cancer 1985;56:531–538.
12. Montie JE, Wood DP, Pontes E, Boyett JM, Levin HS: Adenocarcinoma of the prostate in cystoprostatectomy specimens removed for bladder cancer. Cancer 1989;63:381–385.
13. Kahn P: Coming to grips with genes and risk. Science 1996;274: 496–498.
14. Johnson FE, Virgo KS, Edge SB, et al. (eds): ''Cancer Patient Follow-up.'' St. Louis: Mosby, 1997.